3a - Browsing SNP Data by Association Finding

3a - Browsing SNP Data by Association Finding

You can search the caGWAS database and report for specific genes or regions of interest. This chapter describes **Association finding**, to identify regions of chromosomes associated with cancer.

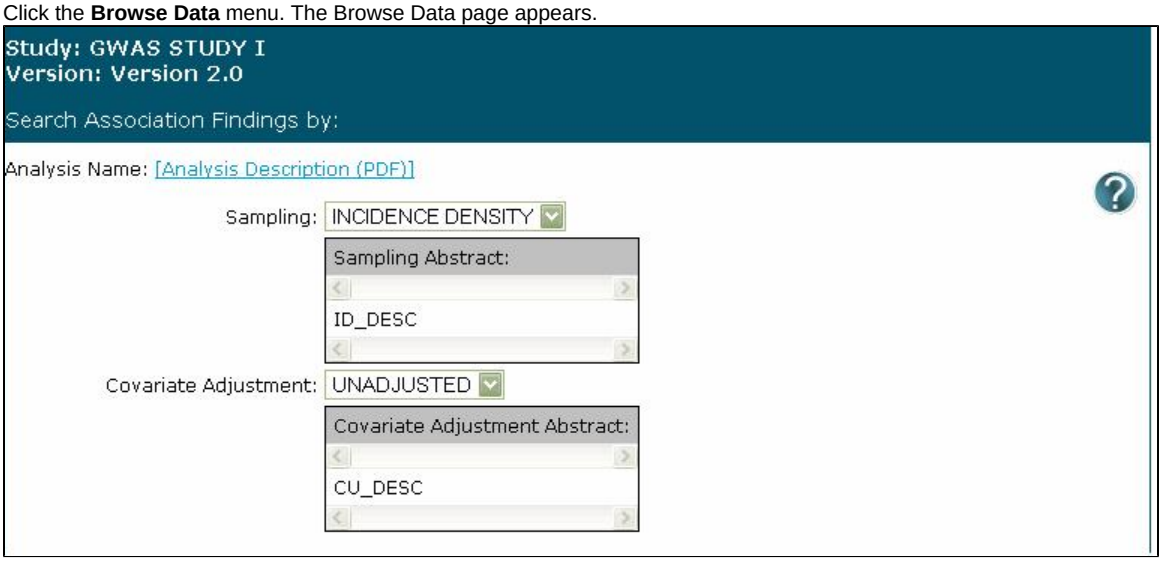
This chapter includes the following topics.

Searching SNP Data by Association Finding

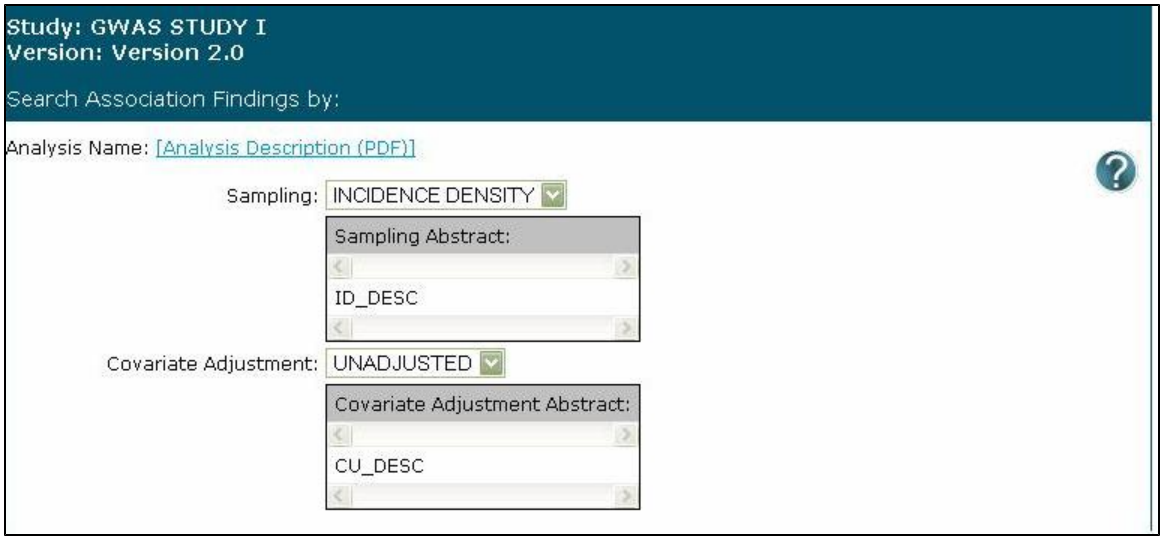
This type of search identifies regions of chromosomes associated with cancer. For each study in the caGWAS data portal, you may refine your search by selecting analysis categories uniquely relevant to the study.

To search by association finding

1. Click the **Browse Data** menu. The Browse Data page appears.



2. Select a study and a version. If you do not select a version, your search returns the most recent study data.
3. Click the **Association Finding** dataset option.
4. Click **Submit**. The Search Association Findings page appears.



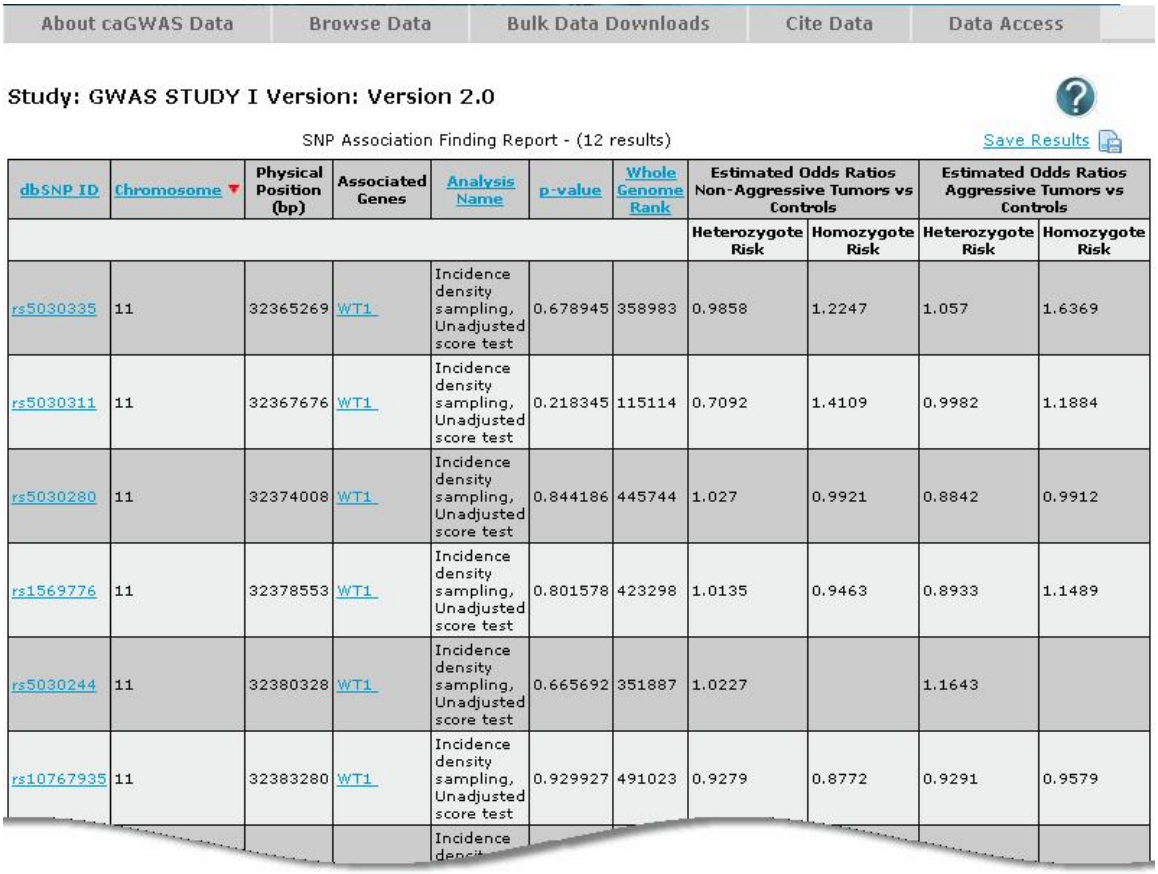
Note
All of the following search criteria are optional. The more you refine your search by adding search criteria, the fewer results you receive. On the Browse Data page you must select at least one search criterion to filter your results. If you want to see the entire dataset, select the Bulk Data Download menu instead. For more information, see 4 - Downloading Bulk Data.

5. In the **Analysis Name** section, select an analysis category from each of the available drop-down lists. After you select a category, a definition of that category appears below the list. For more information about the analysis categories, click the **Analysis Description (PDF)** link at the top of the page.
6. To further refine your search, specify any combination of the following criteria.

Search Criteria	Explanation
Panel	Select a SNP panel.
Genomic Location	To search by genomic location, select a chromosome number from the drop-down menu. Specify a specific chromosome region in the **From** and **To** boxes.
HUGO Gene Symbol	To search by gene symbol, enter one or more HUGO gene symbols (for example, MET) in the text block, or, to upload a list of up to 100 gene symbols, click **Browse** to select the file.
dbSNP Identifier	To search by dbSNP ID, enter one or more dbSNP IDs (for example, rs38840), or, to upload a list of up to 1000 dbSNP IDs, click **Browse**.
p-value	To search by p-value, select **p-value** to enter a value between 0 to 1.
Whole Genome Rank	To search by rank of as so cation, select **Whole Genome Rank** and enter an integer that represents a cut-off rank.

Note
To clear your selections and start a new search, click **Reset**.

7. Click **Submit.** The SNP Association Findings Report appears.



dbSNP ID	Chromosome	Physical Position (bp)	Associated Genes	Analysis Name	p-value	Whole Genome Rank	Estimated Odds Ratios Non-Aggressive Tumors vs Controls		Estimated Odds Ratios Aggressive Tumors vs Controls	
							Heterozygote Risk	Homozygote Risk	Heterozygote Risk	Homozygote Risk
rs5030335	11	32365269	WT1	Incidence density sampling, Unadjusted score test	0.678945	358983	0.9858	1.2247	1.057	1.6369
rs5030311	11	32367676	WT1	Incidence density sampling, Unadjusted score test	0.218345	115114	0.7092	1.4109	0.9982	1.1884
rs5030280	11	32374008	WT1	Incidence density sampling, Unadjusted score test	0.844186	445744	1.027	0.9921	0.8842	0.9912
rs1569776	11	32378553	WT1	Incidence density sampling, Unadjusted score test	0.801578	423298	1.0135	0.9463	0.8933	1.1489
rs5030244	11	32380328	WT1	Incidence density sampling, Unadjusted score test	0.665692	351887	1.0227		1.1643	
rs10767935	11	32383280	WT1	Incidence density sampling, Unadjusted score test	0.929927	491023	0.9279	0.8772	0.9291	0.9579
				Incidence						

8. If your query results in more than 500 records, caGWAS generates a report and sends it to you by email. For more information, see Build and Deployment Automation Exceeding 500 records Note.

Understanding the SNP Association Findings Report

> **Note**
> If your report exceeds 500 records, you do not receive that report immediately. caGWAS prepares a report, and emails it to you. caGWAS informs you of the size of the report you requested, and prompts you to enter your email address. Click **Submit**. You will receive an email when your report is available to download. You will have five days to download the report.

The SNP Association Findings report provides information on SNPs analyzed in a genome-wide analysis, indicating the degree of association found between SNPs and phenotypes.

Explanations of the columns in the SNP Association Findings report appear in the following table.

Column Number	Column	Explanation
1	dbSNP Identifier	SNP identifier, see http://www.ncbi.nlm.nih.gov/SNP
2	Chromosome	Click the column name to sort the report by chromosome and physical location.
3	Physical Position	Position of the SNP on the chromosome in base pairs
4	Associated Genes	Genes within 15,000 base pairs upstream and 10,000 base pairs downstream of the SNP
5	Analysis Name	• Incidence density sampling, Unadjusted score test or • Incidence density sampling, Adjusted score test
6	p-value	Significance value of the analysis - a lower value means the result is more significant

7	Whole Genome	Rank of the SNP's p-value compared to all other SNPs in the genome scan (higher rank number indicates lower significance)
8	Estimated Odds Ratios Non-Aggressive Tumors vs. Controls	Estimated odds ratios for non-aggressive tumors vs. controls for heterozygote and homozygote risk
9	Estimated Odds Ratios Aggressive Tumors vs. Controls	Estimated odds ratios for aggressive tumors vs. controls for heterozygote and homozygote risk

Notes about SNPs

The SNPs are in order of p-value for the unadjusted analysis.

- A SNP that is not close to a gene (as defined in column 4 above) will not have a value for "Associated Genes".
- There are two rows for each SNP - one for the adjusted score, and one for the unadjusted score.
- Description of analyses
  - Unadjusted: A 3-by-m contingency table of genotypes by phenotypes was constructed.
  - Adjusted: The m phenotypes were regressed on indicator variables for genotype effects, age group at randomization (4 groups), region of recruitment (9 non-reference regions), and an indicator variable for cases diagnosed within one year of entry to the trial, and 3 sets of eigenvectors from the most significant principal components.

You can click a column name link to sort the whole report by that column. Sorting order toggles between ascending order (low to high) and descending order (high to low) by repeated clicks. An arrow image indicates the sorted order of the column as or .

To save the data to a tab-delimited text file, click **Save Rsults**. Your browser prompts you to open or save the file. For convenient data viewing, save the file and open it in a spreadsheet application.

3 - Browsing Data in caGWAS

3b - Browsing SNP Data by Population Frequency